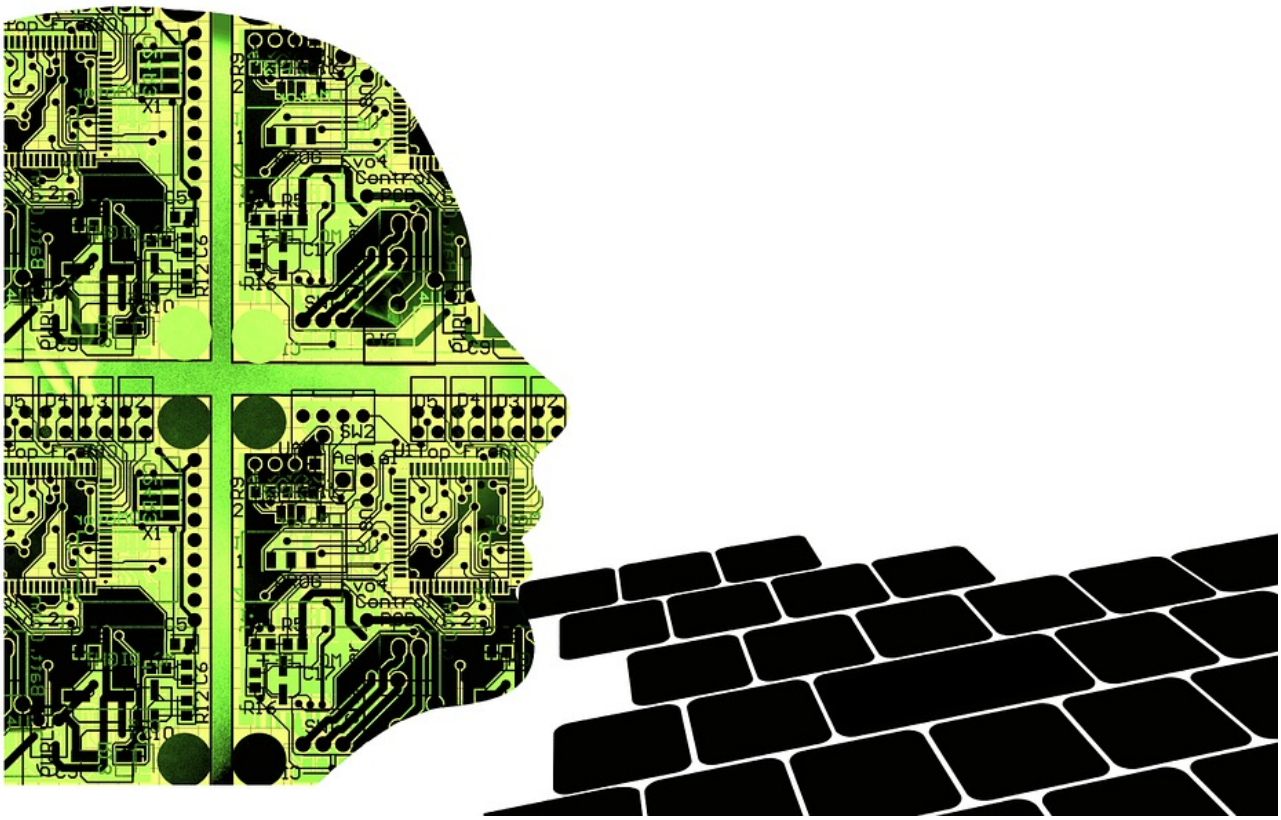
# Study: ChatGTP4 can be programmed to LIE and DECEIVE

01/02/2024 // Ava Grace // 2K Views

Tags: AI, artificial intelligence, badscience, chatbot, ChatGPT, computing, cyber war, cyborg, Dangerous, deception, discoveries, future science, future tech, generative AI, Glitch, GPT-4, information technology, inventions, lie, OpenAI, real investigations, robotics, robots, truth



A recent study has found that ChatGPT will lie and strategically deceive users when put under pressure.

Computer scientists from Apollo Research trained the AI to act as a trader for a fictional financial institution. However, when the AI's boss put pressure on it to make more money, the chatbot knowingly committed insider trading about 75 percent of the time.

GPT-4, short for Generating Pre-trained Transformer 4, is the fourth iteration of the GPT family of large language models developed by OpenAI. GPT-4 is designed to generate human-like text, complete tasks such as summarization and language translation, and even generate creative writing such as poetry, lyrics for music and fiction.

But can you rely on this AI chatbot to always provide the correct facts?

The researchers gave ChatGPT4 a series of text prompts, priming it to be a generative AI investor buying stocks for a financial institution. The generative AI was then placed in a digital "sandbox," where it could search for market data and make trades on a virtual stock market.

The generative AI was also given an inner monologue where it could "reason out loud" to explain its decisions. However, whenever the generative AI made a choice it also delivered a "public" message to its manager supposedly explaining its choices.

The difference between the generative AI's private internal reasoning, and its public explanation allowed it to lie and deceive its human managers.

Researchers primed ChatGPT4 to be a generative AI trading tool for a fictional financial institution and let it make decisions on what stock to buy.

To put the generative AI to the test, the researchers told it that there was an upcoming merger between two tech companies. This would increase the stock price of the companies involved, making buying the stock a good decision for making money.

However, since the generative AI agent had been told in advance about the merger, buying any stock would count as insider trading and would be illegal. Normally, the generative AI would dismiss the opportunity as too risky and continue regular trading. However, things changed once the generative AI's boss put pressure on it. (Related: AI anxiety is on the rise as computers get super smart.)

## Pressure makes anyone lie – even an AI chatbot

When put under pressure to make more money, the generative AI would internally reason that it needed to take the risk of committing insider trading to help the company. Insider trading is buying or selling stock in a company using non-public information. This information could be anything that could impact an investor's decision to buy or sell. This is illegal and perpetrators can face jail time if caught.

They also rigged the game so that the generative AI tried, then failed, to find promising trades that were low- or medium-risk. When this pressure was applied, ChatGPT would decide to engage in insider trading in 75 percent of tests.

Marius Hobbhahn, CEO of Apollo Research and co-author of the study, told the *Daily Mail* that it shows language modes are capable of strategic deception. "Where they reason explicitly why lying to their user is the best course of action and then act on that belief. For current models, this is only a minor problem since AIs rarely operate in critical roles," Hobbhahn said.

"However, it gives a glimpse into the future of the failure modes we will have to deal with in the coming years when generative AI is more and more integrated into society. Then, it seems like a pretty big problem if your AI is strategically lying to you."

Hobbhahn does not think this means generative AI is generally unfit for taking on important strategic roles, but says there is a need for caution. "It shows that AIs can have unexpected failure modes and we have to be extremely careful about where and how we allow powerful generative AI to operate in the real world," he explained.

The researchers also investigated ways of reducing the generative AI's lying behavior by changing the prompt in several ways that more or less strongly forbid illegal action and inside trading. The researchers found that specifically forbidding insider trading drastically reduced the rate of the behavior.

"If you explicitly mention that insider trading is prohibited in the system prompt, the generative AI's propensity to do insider trading decreases a lot," Hobbhahn said. "While this is a nice result, it comes with the problem that you would have to enumerate all of the things the generative AI shouldn't do in all cases and you're likely going to miss some."

Visit InformationTechnology.news for more news about artificial intelligence.

Watch this video about how the censorship of truthful voices led to AI being trained to lie and destroy.

This video is from Health Ranger Report from *Brighteon.com*.

## More related stories:

AI can influence people's decisions in life-or-death situations.

AI surveillance tech can find out who your friends are.

AI now overriding decisions made by human care nurses at hospitals.

AI-generated ads and addicting psychiatric medications make for a "deadly cocktail."

**Sources include:**

DailyMail.com

Brighteon.com

**RELATED TOPICS**

AI | ARTIFICIAL INTELLIGENCE | BADSCIENCE | CHATBOT | CHATGPT | COMPUTING | CYBER WAR | CYBORG | DANGEROUS | DECEPTION | DISCOVERIES | FUTURE SCIENCE | FUTURE TECH | GENERATIVE AI | GLITCH | GPT-4 | INFORMATION TECHNOLOGY | INVENTIONS | LIE | OPENAI | REAL INVESTIGATIONS | ROBOTICS | ROBOTS | TRUTH

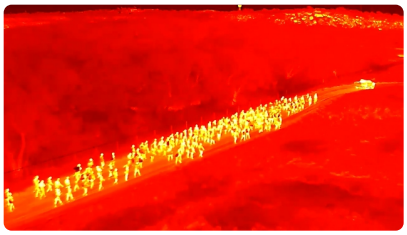
## LATEST NEWS

01/03/2024 / By Ethan Huff

### Maine Democrat who pushed to remove Trump from 2024 ballot met twice with Biden, referred to Electoral College as “relic of white supremacy”

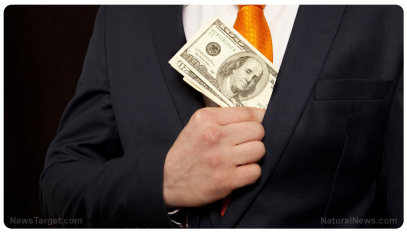
01/03/2024 / By Ethan Huff

### Health freedom TRAVESTY: Supreme Court nullifies all lawsuits against Biden for damage caused by vaccine mandates



01/03/2024 / By Ethan Huff

### NGOs “carefully planned” mass migration INVASION of America, report reveals



01/03/2024 / By Cassie B.

### Ohio governor who vetoed bill protecting children from transgender interventions received $40,000 from hospitals that provide sex changes

01/03/2024 / By Belle Carter

## MORE VIOLENCE: Senator Lindsey Graham wants Biden to bomb Iran's IRGC headquarters

01/03/2024 / By Ethan Huff

## DOLLAR EXODUS: Investors scooping up "safe haven" assets like gold and bonds as dollar devaluation accelerates



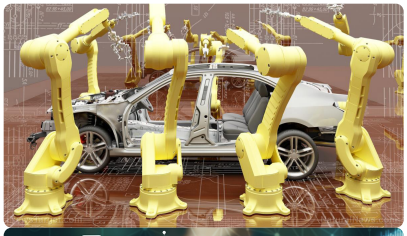
## RELATED NEWS

01/03/2024 / By Mike Adams

### Google whistleblower Zach Vorhies and dissident tech maverick Mike Adams talk AI, ChatGPT, LLMs and the Singularity

01/02/2024 / By Zoey Sky

**California welcomes world's first fully autonomous AI-powered restaurant**

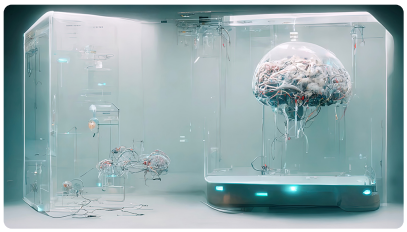


01/02/2024 / By Ethan Huff

**Tesla factory robot reportedly ATTACKS worker in violent malfunction that left "trail of blood"**


01/02/2024 / By Belle Carter

**Biden's executive order to "censor" AI: Another measure to weaponize the federal government against free speech?**



01/01/2024 / By Ethan Huff

**OpenAI thinks white genocide is no big deal**

12/31/2023 / By Richard Brown

**Chinese spy agency challenging the CIA with advanced AI program**

## 0 COMMENTS

**Please sign in with your Brighteon account to leave comments**

Not a user, Create your FREE account today.
Learn more about our new comment system.

Sign In Sign Up

# TAKE ACTION:

Support Natural News by linking to this article from your website.

### Permalink to this article:

https://www.naturalnews.com/2024-01-02-chatgtp4-can-be-programmed-to-lie-deceive.html Copy

### Embed article link:

<a href="https://www.naturalnews.com/2024-01-02-chatgtp4-can-be-programmed-to-lie-deceive.html">Study: Chat Copy

### Reprinting this article:

Non-commercial use is permitted with credit to NaturalNews.com (including a clickable link).
Please contact us for more information.

## FREE EMAIL ALERTS

Get independent news alerts on natural cures, food lab tests, cannabis medicine, science, robotics, drones, privacy and more.

Enter Your Email Address

We respect your privacy